

**WL-TR-96-4062**



**MEMORY-BASED COMPUTATIONAL  
INTELLIGENCE FOR MATERIALS  
PROCESSING AND DESIGN**

Y.H. PAO

Case Western Reserve University  
Electrical Engineering  
Cleveland OH 44106-7221

**FINAL REPORT FOR APRIL 1996**

**Approved for public release; distribution unlimited**

**MATERIALS DIRECTORATE  
WRIGHT LABORATORY  
AIR FORCE MATERIEL COMMAND  
WRIGHT-PATTERSON AIR FORCE BASE, OH 45433-7734**

19960621 070

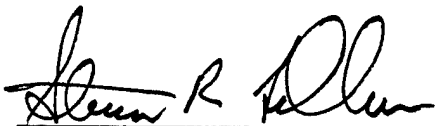
DTIC QUALITY INSPECTED 1

## NOTICE

When Government drawings, specifications, or other data are used for any purpose other than in connection with a definitely Government-related procurement operation, the United States Government incurs no responsibility or any obligation whatsoever. The fact that the government may have formulated or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication, or otherwise in any manner construed, as licensing the holder, or any other person or corporation; or as conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

This report has been reviewed by the Office of Public Affairs (ASC/PA) and is releasable to the National Technical Information Service (NTIS). At NITS, it will be available to the general public, including foreign nations.

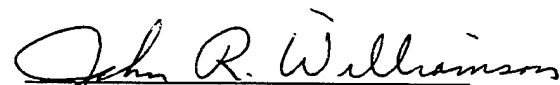
This technical report has been reviewed and is approved for publication.



DR. STEVEN R. LECLAIR, Chief  
Materials Process Design Branch  
Integration and Operations Division  
Materials Directorate



DR. STEVEN R. LECLAIR, Chief  
Materials Process Design Branch  
Integration and Operations Division  
Materials Directorate



JOHN R. WILLIAMSON, Chief  
Integration and Operations Division  
Materials Directorate

If your address has changed, if you wish to be removed from our mailing list, or if the addressee is no longer employed by your organization please notify WL/MLIM Bldg 653, 2977 P Street, Suite 13, WPAFB, OH 45433-7746 to help us maintain a current mailing list.

Copies of this report should not be returned unless return is required by security considerations, contractual obligations or notice on a specific document.

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 0704-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.				
1. AGENCY USE ONLY (Leave blank)		2. REPORT DATE APRIL 1996		3. REPORT TYPE AND DATES COVERED FINAL 12/01/92--12/01/95
4. TITLE AND SUBTITLE  MEMORY-BASED COMPUTATIONAL INTELLIGENCE FOR MATERIALS PROCESSING AND DESIGN			5. FUNDING NUMBERS  c: F33615-87-C-5250 PE 62102 F PR 2306 TA P9 WU 03	
6. AUTHOR(S)  PAO, Y.H.				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)  Case Western Reserve University Electrical Engineering Cleveland OH 44106-7221			8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)  Materials Directorate Wright Laboratory Air Force Materiel Command Wright Patterson Air Force Base, Ohio 45433-7734			10. SPONSORING / MONITORING AGENCY REPORT NUMBER  WL-TR-96-4062	
11. SUPPLEMENTARY NOTES				
12a. DISTRIBUTION / AVAILABILITY STATEMENT  Approved for Public Release; Distribution is Unlimited			12b. DISTRIBUTION CODE	
13. ABSTRACT (Maximum 200 words)  The work reported in this document is concerned with the efficient use of computers in materials research and in applications of the results of that research. Emphasis is on the development of computational methodologies which can facilitate the innovative design of materials and of materials processing, for high performance materials and for composite materials structures. Basic advances have been made in three areas of adaptive computing: in establishing the practice of functional-link neural-net computing for learning models of material behavior, in developing a parallel processing evolutionary search paradigm for optimization, and in exploring various ways of establishing and using associative memories. This report also describes how these innovations in conceptual and theoretical matters can be used to effect in dealing with materials processing and design tasks. Applications include in-situ - real-time interpretation of ellipsometry data, optimal formulation of material composition, control of nonlinear dynamic systems, memory-based design and the implementation of visual displays for multi-dimensional data.				
14. SUBJECT TERMS Neural Networks Evolutionary Programming			Product Formulation Optimization	
			Process Monitoring Intelligent Control	
			15. NUMBER OF PAGES 60	
			16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT UNCLASSIFIED	18. SECURITY CLASSIFICATION OF THIS PAGE UNCLASSIFIED	19. SECURITY CLASSIFICATION OF ABSTRACT UNCLASSIFIED	20. LIMITATION OF ABSTRACT  SAR	

## TABLE OF CONTENTS

SECTION	PAGE
1. INTRODUCTION TO REPORT.....	1
2. THE FUNCTIONAL-LINK CONCEPT IN NEURAL NETS.....	6
The Basic Idea.....	6
Relationship to the Kolmogorov Superposition Theorem and to Basis Function Expansions.....	9
3. APPLICATIONS OF THE FUNCTIONAL-LINK APPROACH.....	16
Inverting The Fresnel Equations for Interpreting Ellipsometry Data.....	16
System Identification and Noise Cancellation.....	20
4. GUIDED EVOLUTIONARY PROGRAMMING.....	22
Overview of Optimization Algorithms.....	22
Guided Evolutionary Programming with Simulated Annealing.....	26
Applications Including The Interpretation of Ellipsometry Data.....	27
5. APPLICATION OF FUNCTIONAL-LINK AND GUIDED EVOLUTIONARY PROGRAMMING TO MATERIALS FORMULATION.....	32
6. APPLICATION OF FUNCTIONAL-LINK AND GUIDED EVOLUTIONARY PROGRAMMING TO OPTIMAL CONTROL OF NONLINEAR DYNAMIC SYSTEMS.....	34
7. THE ROLE OF SELF-ORGANIZATION IN DATA ANALYSIS: Use in The Efficient Storage and Retrieval of Design Information.....	37
8. REDUCED DIMENSION REPRESENTATION OF DATA: The Constrained Variance Approach.....	39
9. SUMMARIZING REMARKS.....	42
10. REFERENCES.....	44
11. ATTACHMENTS: REPRINTS, PREPRINTS AND DISKETTES.....	46

## LIST OF ILLUSTRATIONS

Figure 1.	Single Hidden-Layer Feedforward Net.....	7
Figure 2.	Functional-Link Net with Enhancement Nodes.....	7
Figure 3.	The original Komogorov Theorem.....	13
Figure 4.	Single Hidden Layer Feedforward Net.....	14
Figure 5.	Basis Function Expansion and Functional-Link.....	15
Figure 6.	Illustration of Reflection from a Composite Structure.....	18
Figure 7.	Input and Output Parameters Used in Neural Net.....	19
Figure 8.	Neural Net for Inverting Fresnel Equations.....	20
Figure 9.	A Skeletal Form of The Basic GA/EP Algorithm.....	23
Figure 10.	The Basic SA Algorithm.....	24
Figure 11.	The GESA Algorithm.....	28
Figure 12.	Noisy Ellipsometry Data.....	30
Figure 13.	The Modeller/Predictor Neural Net.....	35
Figure 14.	The Neural Control-Action Generator.....	36

## LIST OF TABLES

Table 1.	Estimates of Pseudosubstrate and Film Optical Constants.....	31
----------	--------------------------------------------------------------	----

## FOREWORD

This report was prepared by Case Western Reserve University (CWRU), Cleveland, Ohio, under USAF Contract No. F33615-87-C-5250. This is the final report summarizing the results of research carried out over a period of seven years from February 1988 to December 1992. This report covers work carried out by CWRU faculty, staff, and students. However, this work was carried out in close co-ordination with researchers in the Process Design program of the Materials Directorate of Wright Laboratory, with some of the work done on-site at WPAFB. Work was administered by Dr. Steven R. LeClair who also contributed to the themes upon which the research was based. Although a final report for the entire project, this document nevertheless concentrates primarily on work done during the period October 1992 to December 1995. Results obtained prior to that have been reported in an interim report WL-TR-93-4021. All along, results from this program have also been mentioned in technical reports generated by Dr. LeClair, and have been published in technical journal articles. This report is in the nature of a collection of very brief discussions each describing an issue and the need for achieving an improvement in the matter. The progress achieved in each case and the practical significance of the advance are also described briefly. In all cases, details are made available through attached reprints of published technical journal articles or with the use of software diskettes.

Contributors to the results reported in this documents include the following:

Faculty: Yoh-Han Pao (P. I.)

Frank L. Merat

Senior Research Associates: Dejan J. Sobajic,

Boris Igelnik

Research Associates:	Percy P. C. Yip Wassim Hafez
Research Assistant:	Michal S. Soclof
Visiting Researcher:	Professor C. Y. Shen
Graduate Students:	Thomas Hemminger, R. B. Cai, G. H. Park, Z. Meng, P. P. C. Yip, David Yan



## **ACKNOWLEDGMENTS**

The insightful strategic guidance of Dr. Steven R. LeClair over the span of this multi-year contract is acknowledged with deep appreciation. It is also a pleasure to acknowledge the strong collaborative support of researchers at Wright Laboratories, and also at AI Ware Inc., in the various efforts described in this document.

It is a duty as well as a pleasure to acknowledge the beneficial effects of collaborations with Professor Phil Chen of Wright State University in the matters of adaptive learnings and several other topics related to intelligent planning in manufacturing.

The help of Mr. Ron Cass and Mr. Ed Wilson of AI Ware Inc. in technical matters and in making the CAD/Chem diskettes available is gratefully acknowledged.

## 1. INTRODUCTION TO REPORT

This document reports on the results of research carried out over a number of years by Case Western Reserve University faculty and students in the area of adaptive distributed parallel processing in support of materials research, in collaboration with the Materials Process and Design Research group of the Materials Research Directorate of Wright Laboratory. The work was carried out over an extended period of time, ranging from February 1988 to December 1995. However this document concentrates on the period from October 1992 to December 1995. Previous results had been reported in report WL-TR-93-3021.

The keynote of the work was the development of adaptive and self-directed computational methods to be used in support of materials research so that the efficiency of materials research could be improved significantly.

It so happens that dramatic advances were being made in the development of neural-net computing, evolutionary programming, and other adaptive parallel computational methodologies during this span of time. This present research effort was able to participate and contribute significantly to the overall flow of events, and this story is told and documented with this report.

The story is multi-faceted and has many inter-related parts. But the essential fact is that this program was able to make two major useful contributions to the practice of adaptive parallel computing, and also worked to show how these contributions can be used to increase the efficiency of materials research.

One of the two advances is in the area of supervised learning or the training of neural-nets for pattern recognition and for function approximation. The idea of use of

'functional-links' turned out to be a powerful and liberating one, and practice of that approach has been of great help in the work reported in this document. The so-called radial basis function approach is but a specialized instance of similar practice.

Another basic advance is the development of a parallel evolutionary stochastic search method which also uses simulated annealing to avoid local minima.

These two advances can be used in combination, for example, to model a technology and to point to optimal operating conditions, or to discover optimal material formulations. This is described in this report and demonstrated with the help of a software diskette.

The functional-link approach to learning can dramatically shorten the time required to train nets, so much so that rather complicated tasks become feasible. Some such applications are described in this document. One application is the use of the functional-link for the rapid in-situ interpretation of experimental parameter readings in ellipsometry monitoring of molecular beam growth of thin film structures. In another instance, the ease of retraining facilitates process monitoring and control tasks; changes in the nature of a process, or in the response of a sensor, or in the action of a transducer can be detected and modeled adaptively in real time.

The functional-link approach is discussed briefly in Section 2 of this report, and the discussion is substantiated with inclusion of two reprints of technical journal articles. The point made in one of the articles is that if the functional-links are volunteered, rather than learned, then the network learning task can become a linear one, handled efficiently with methods such as conjugate gradient search, in a small number of steps. Both the learning and generalization characteristics of the functional-link net are very good, often much superior to those of the conventional multilayered net trained with the Backpropagation

algorithm. The point of the other article is that the functional-link approach can be given a rigorous mathematical base.

There is no question that the generalized Perceptron nodal architecture of such nets is responsible for their representation and learning power . There is also no doubt that the use of the Perceptron architecture was 'a leap of faith' inspired by the results of neuroscience research. A preprint of a technical article entitled 'A Historical Perspective on Neural-Net Computing' is attached to this document .

An application of the functional-link net is discussed briefly in Section 3 of this report. The task is the inversion of Fresnel equations for the purposes of estimating the values of optical constants and thickness of thin films. Details are provided in a reprint and a preprint of technical journal articles made available as attachments to this document.

Another contribution of this research program is the development of a parallel version of stochastic search in optimization computations. This is explained briefly in Section 4 with the help of an illustration of how the Fresnel equations might also be inverted with evolutionary programming. The innovation in this new algorithm is the use of several intercommunicating searches proceeding concurrently in parallel, providing guidance to each other. Simulated annealing is used to avoid local minima as much as possible. Other details are provided in two reprints attached to this report.

The functional-link and guided evolutionary programming technologies can be used to great effect in combination. A powerful and versatile software package used for materials formulation and for the design of experiments is based on these practices and is commercially available. It is described briefly in Section 5 and a demonstration copy of that software, the CAD/Chem system of AI Ware Inc., is appended to this document,

together with instructions for use. It is but a demonstration copy. [CAD/Chem is a trademark of AI Ware Inc.].

It is stated in Section 6 that there are perhaps three distinct manners in which neural-net computing can be used for the control of nonlinear dynamic systems, these being the Backpropagation in time method, the inverse net approach and the optimal control with the feedforward net and optimization. These are described briefly in Section 6 and the third method is described in some detail in an attached reprint.

Perhaps the most basic task of data analysis is that of density estimation, namely finding out where the data are and finding a way of describing the density of such occurrences analytically. Ultimately, it is necessary that such a task be carried out automatically, through self organization. This task is discussed briefly in Section 7. It is as yet an unsolved problem. In the meantime, one form of self-organization is through clustering, and associative memories for materials data and part designs can be implemented with use of hierarchical structures of linked clusters. An indication of what might be done with such cluster based associative memories is provided through discussion of a multimedia associative memory for trouble shooting defects in metal-cast parts. A reprint of a conference proceedings paper and a software diskette are provided for that purpose.

Visualization and display are discussed briefly in Section 8. It is difficult for humans to grasp the significance of a body of multi-dimensional data and presentations of that same data in some reduced-dimension form is often very helpful. A new dimension-reduction method, called the variance constraint method is described briefly in Section 8 together with suggestions of use in support of process monitoring and control. A preprint of a journal article is provided as an attachment to this report. Summarizing remarks are contained in Section 9, and references are listed in Section 10.

The work of this program has influenced the work of others beneficially. One formal indication of that influence is the issuance of one patent to Wright Laboratory and the filing of a second patent application by Wright Laboratory. A third is being evaluated for filing. Material related to the patent and patent application are available in an attachment to this report.

This report is presented in the form of two volumes. Volume 1 contains the text of all the Sections. All the attachments and the two diskettes are contained in Volume 2.

The titles of all attachments are listed in Section 11.

## 2. THE FUNCTIONAL-LINK CONCEPT IN NEURAL-NET COMPUTING

### The Basic Idea

The single most valuable contribution of neural-net computing to the art of computing is undoubtedly the introduction of the practice of function approximation with use of the multi-layer feed-forward net, made up of Perceptron-like nodes. The result is that a multi-variate function is described simply as a single-variable nonlinear transform of the sum of many single-variate functions which in turn are single-variable nonlinear transforms of sums of many single-variate functions and so on, recursively, as much in depth as required.

Also, without a doubt, use of that type of nodal network architecture was inspired and motivated by the results of perhaps 200 years of research in the neuroscience. An account of the development of the use of the multi-layer feed-forward nodal net for various pattern recognition and function approximation purposes is given in Attachment 1, a preprint of an article by Yoh-Han Pao on 'A Historical Perspective on Some Aspects Of Neural-net Computing'. To date that article has served as the basis for tutorials at conferences.

From 1986 onwards, after the publications of Rumelhart et al [1] and others had popularized the practice of the Backpropagation algorithm for learning computational models of functions, and after a multitude of practitioners had indeed experienced for themselves the efficiency of such nets, it was thought that there was something very specific about the multi-layer nature of the net, and that computational power would be lost if significant changes were introduced. It was in that environment and under those circumstances that Pao and his collaborators [2] introduced the concept of the functional-link net.

Briefly, the suggestion was made that instead of always using a multi-layer architecture and the Backpropagation-of-error algorithm for learning all network parameters, one might volunteer some nonlinear functional transforms to circumvent part of the tiresome iterative parameter adjustment procedure. Thus even a single hidden layer net, such as that shown in Figure 1, might be simplified through use of appropriate functional links, as shown in Figure 2.

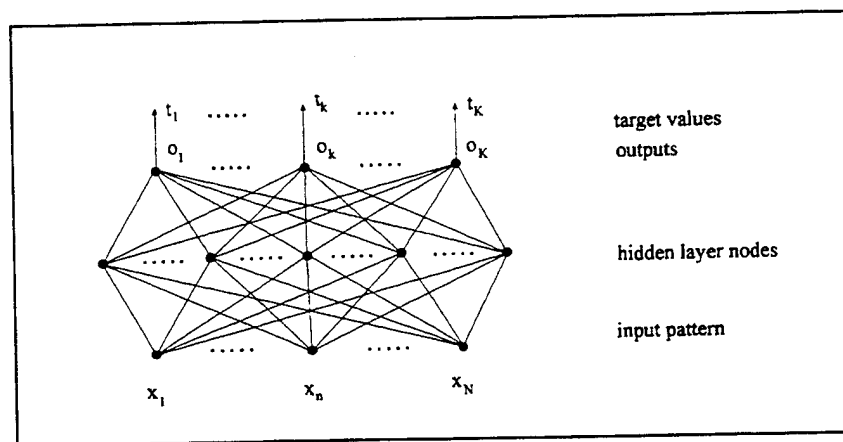


Figure 1 Single Hidden-Layer Feedforward Net.

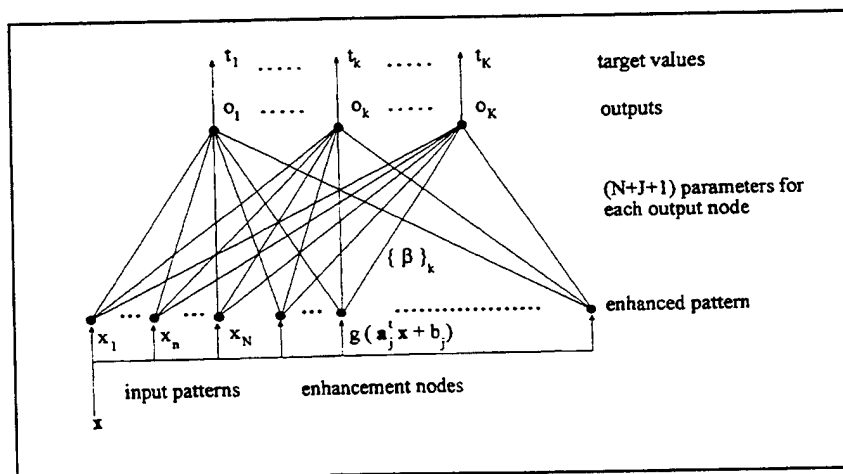


Figure 2 Functional-Link Net with Enhancement Nodes.



The critical point is that parameters in the functional-links do not have to be learned iteratively and laboriously. As illustrated in Figure 2, the learning task becomes a linear one, and elegant and efficient algorithms such as that of conjugate gradient search become valid for use.

It turns out that the functional links might even be chosen 'randomly' within constraints. An account of a 'random vector' version of the functional-link approach is contained in Attachment 2, a reprint of a Neurocomputing publication by Pao et al [3] on 'The Learning and Generalization Characteristics of the Random Vector Functional-link Net'. In that article, it is shown that a functional link net can be trained very rapidly and can have good generalization characteristics. Subsequently Igel'nik and Pao [4] provided a rigorous theoretical basis for that approach, as described in Attachment 3, a reprint of an IEEE Transactions paper entitled 'Stochastic Basis Functions and the Functional-link Net'.

The concept of the Functional-link net has been a liberating one, allowing for experimentation outside of the strict practice of the Backpropagation-of-Error algorithm. Depending on the circumstances, the nature of the functional-links could be quite different from one case to another. In some instances, localized functions such as Gaussians would be convenient and appropriate; in other cases, distributed functions would be more efficient and in yet other cases, functions such as wavelets would provide some measure of both localization and distribution. Currently, at the time of preparation of this report, the use of 'radial basis functions' has become very popular, and these are used and misused widely. However, in fact, these can be regarded as one instance of the functional-link approach.

Actually, by themselves, the so-called 'radial basis functions', such as Gaussians, are not legitimate basis functions, and there is no theoretical justification for using a small finite number of such functions for providing a basis for a complex non-linear function. Indeed there would need to be an exponentially large number of such functions to provide an adequate basis for description of a multivariate function. It is the limit integral and the Monte Carlo method for evaluating that integral, as described in Attachment 3, that provides the theoretical justification for such an approach and also provides guidance on how to choose the relatively small finite number of 'basis' functions'.

The Functional-link approach has been used extensively to advantage in the work reported in this document.

#### Relationship to the Kolmogorov Superposition Theorem

In this subsection, we pause to examine whether the supervised learning functionality of neural-net computing is indeed a new contribution to the practice of computing, or whether perhaps it might be some previously known method, renamed.

In the case of supervised learning, the question is whether it is possible to infer the values of a function over a continuous domain, given only values of that function for a discrete set of sample points in that domain. This task might be viewed as reconstruction of a function, or learning a function, or function approximation. The task is very difficult if the function is a multivariate one; but that functionality is very much needed in information processing. It is the essence of modeling, estimation, prediction and other related tasks.

It is known from Shannon's Theorem [5] that a one-dimensional band-limited function can be reconstructed in total over the entire continuous domain from values of the function at a discrete set of sampling points. Extension of Shannon's Theorem to the multidimensional case can be done in a straightforward manner, to yield a procedure which grows exponentially in computational complexity with (linear) increase in the number of dimensions; that is one aspect of what is sometimes referred to as the 'curse of dimensions' [5]. It is to exorcise this curse that it seemed the supervised learning functionality of neural-net computing had to be invented and it was attempted at first in an empirical manner because of inspiration from the neuron doctrine. The present question is whether similar procedures had already been proposed and made available by traditional mathematical methods.

The answer is very interesting. It develops that Kolmogorov [6] and Sprecher [7] had proved remarkable results regarding the representation of continuous functions of several variables. The Kolmogorov Superposition Theorem proved that such multi-variate functions can be expressed as superposition of functions of one variable and by sums of functions.

In particular Kolmogorov proved the following theorem:

Theorem (Kolmogorov). There exist fixed continuous increasing functions  $\psi_{qi}(x)$ , on  $I = [0,1]$  such that each continuous function  $f$  on  $I^s$  can be written in the form

$$f(x_1, \dots, x_s) = \sum_{q=0}^{2s} g_q \left[ \sum_{i=1}^s \psi_{qi}(x_i) \right] \quad (2.1)$$

where the  $g_q$  are properly chosen continuous functions of one variable.

Sprecher showed that the  $\psi_{qi}$  functions could be replaced by  $\lambda_i \phi_q$ , yielding a stronger version of Kolmogorov's theorem:

Theorem (Sprecher). There exists constants  $\lambda_i$  and fixed continuous increasing functions  $\phi_q$  on  $I = [0,1]$  such that each continuous function  $f$  on  $I^s$  can be written in the form

$$f(x_1, \dots, x_s) = \sum_{q=0}^{2s} g_q \left[ \sum_{i=1}^s \lambda_i \phi_q(x_i) \right] \quad (2.2)$$

In this theorem, the  $g_q$  functions depend on  $f$ . The constants  $\lambda_i$ , and the functions  $\phi_q$  do not depend on  $f$ , and are universal functions which can be used for any  $f$  function.

The original Kolmogorov theorem can diagrammed as a feedforward net shown in Figures 3, for comparison with a conventional multilayer feedforward neural net of the Perceptron type, shown in Figure 4. In analogous manner, a functional-link net or basis function net would correspond to the depiction of Figure 5.

The net for the original Kolmogorov result indicates that it is possible to represent a multivariable function as the sum of a finite number of single variable functions, each of which is a function of a sum of a finite number of universal single variable functions, functions which do not vary in form with the task at hand. This is a marvelous result except for the fact that the single variable functional forms are not known and there is no practicable constructive procedure for developing those functions.

It is not known if those mathematics results had any effect on the evolution of neural-net computing prior to the resurgence stage. More recently, Hecht-Nielsen [8] drew attention to the relevance of the Kolmogorov and Sprecher results to neural-net computing, and Sprecher [9] reported on yet another formulation of the Kolmogorov theorem, simplifying it further and bringing it closer to the format of the Perceptron and neural-net architecture.

In retrospect, it must be admitted that the multilayer feedforward net with Perceptron nodes is a genuine original contribution inspired by biology rather than by mathematics. The results of Kolmogorov and Sprecher are reassuring and supportive on the one hand, and are challenging on the other, showing how simple function approximation could be if we only knew the correct functional forms of the various 'basis' functions. Of particular interest is the fact there exist 'universal' functions at the lower level which do not depend on the problem, but can serve for all the function approximation tasks.

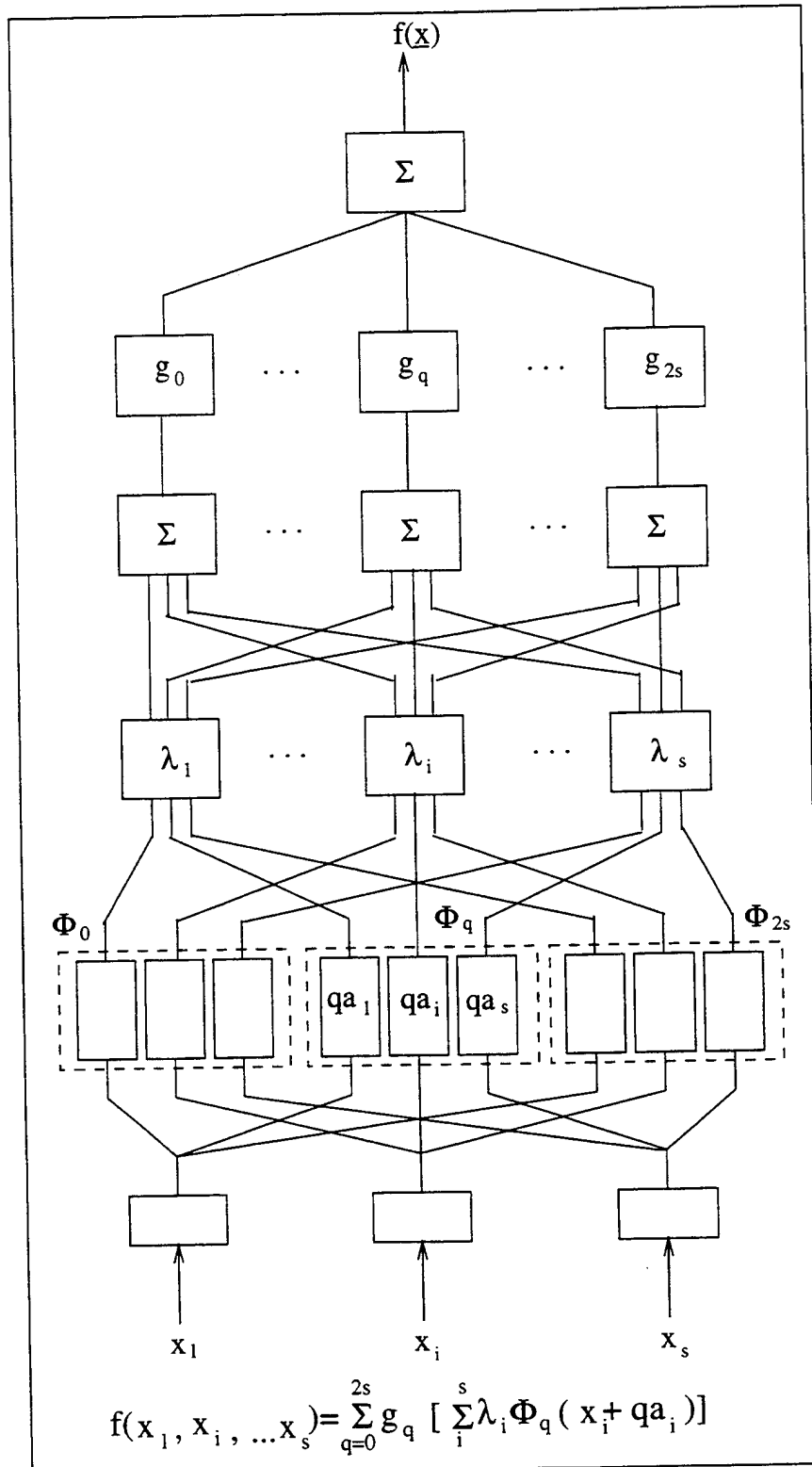


Figure 3 The original Komogorov Theorem.

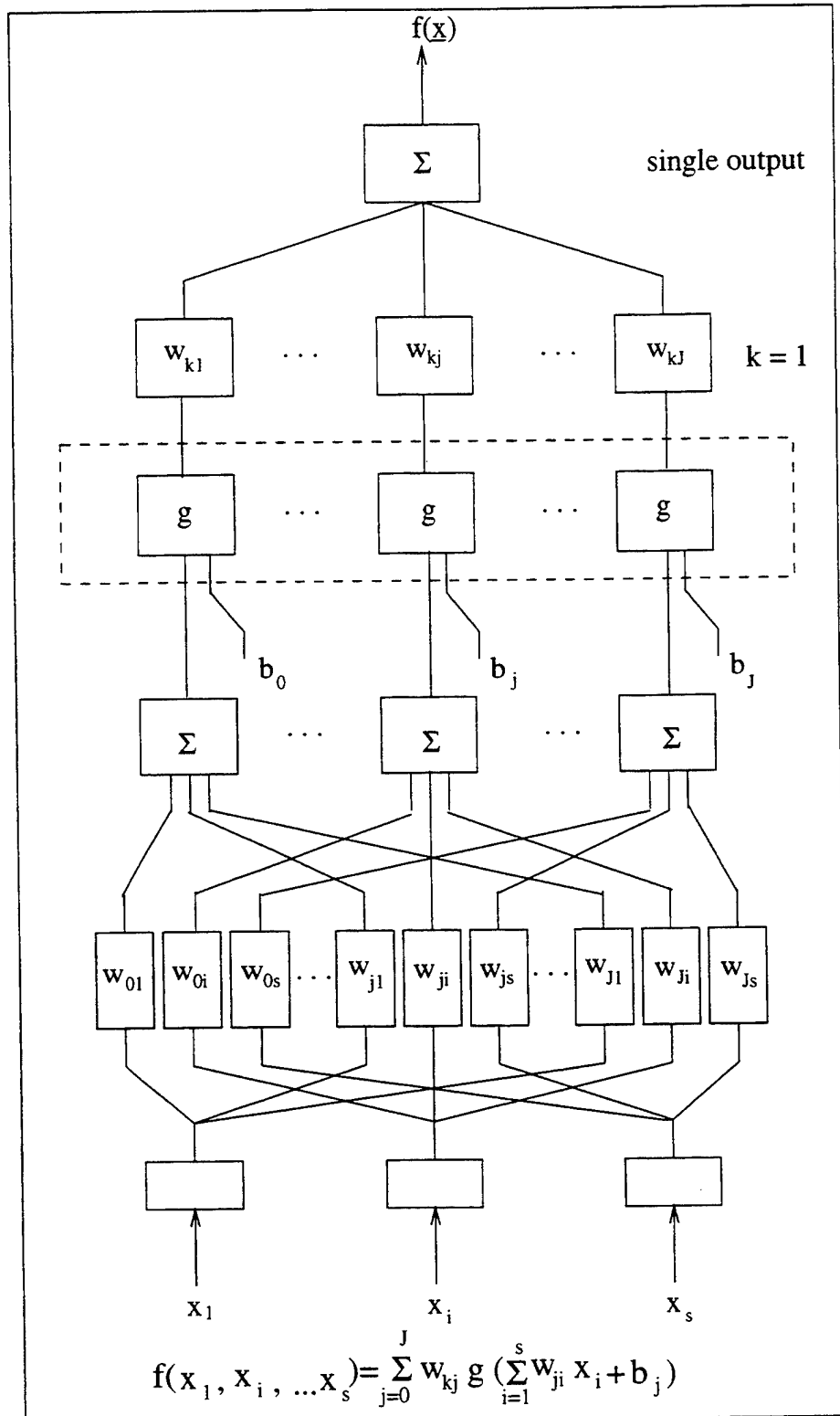


Figure 4 Single Hidden Layer Feedforward Net.

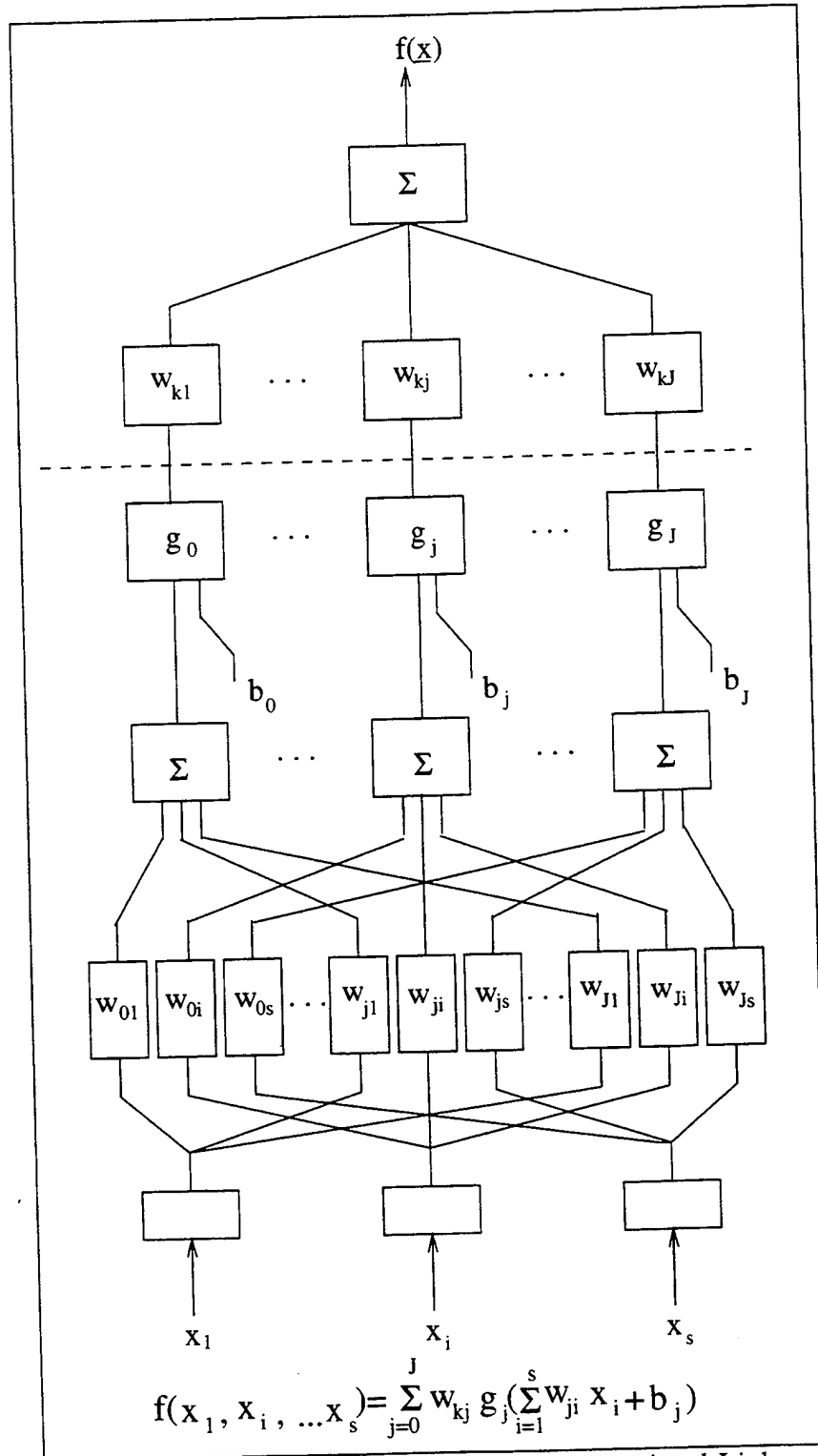


Figure 5 Basis Function Expansion and Functional-Link.



### 3. APPLICATIONS OF THE FUNCTIONAL-LINK APPROACH

#### Parameter Interpretation: Inverting the Fresnel Equations for Interpreting Ellipsometry Data.

The multilayer feedforward neural-net can be used in a number of ways for various applications, one of which is that of process parameter interpretation.

This subsection of this report describes the use of neural-net computing as an enabling factor in a scheme for real time monitoring of the growth of multi-layer thin film structures of semiconductor materials.

The films in question are grown with Molecular Beam Epitaxy (MBE) or with variations on the theme[10]. In such growth, components of the semiconductor components are heated in crucibles so that vapors are produced in the separate chambers. Shutters are opened alternatively in controlled manner to produce molecular beams of the component compositions. Conditions can be found so that thin semiconductor films of the desired stoichiometry are grown on heated substrates in a controlled manner. The films can be of uniform thickness and can be formed in superlattices with the thickness of individual layers varying from about 10 Angstroms to about 3000 Angstroms or so.

In such operations, the crucible temperatures, vapor pressures, shutter timing and substrate temperature all need to be controlled if the desired compositions, film thickness, and physical properties are to be attained. This means, in turn, that accurate in-situ monitoring is essential and it seems that optical ellipsometry is suitable for providing some aspects of that essential in-situ monitoring[11][12]. In particular, the polarization parameters of light reflected from the surface of such film structures can be interpreted to yield information on the complex refractive index of the film being grown. That

particular interpretive action may be achieved in a number of ways but it can also be carried with use of the functional-link net, and that practice is described in this subsection.

Detailed accounts of this approach have been presented at Symposia and have been published in technical journals. A brief technical discussion is contained in Attachment 4, and a preprint of a more detailed IEEE Transactions paper is available as Attachment 5.

For the present qualitative purposes, it suffices to say that when a beam of circularly polarized light is reflected from the surface of a material with a complex refractive index (a partially absorbing material), the reflected beam is observed to be elliptically polarized. This is because the in-plane and out-of-plane components of the incident beam are reflected with different reflectances and an additional phase shift is introduced between the two components. Given the refractive index of the substrate and of the film, and the thickness of the film, it is possible to calculate the ellipticity parameters with use of the Fresnel equations. The question is how to invert the procedure, so as to be able to infer knowledge of the film refractive indices and thickness, given one or more sets of ellipsometry readings.

It is suggested that neural-nets be used for inverting the Fresnel equations. The validity of the procedure can be demonstrated readily with calculated results. Of greater interest is to ascertain how that method functions with noisy data.

The experimental situation is illustrated schematically in Figure 6.

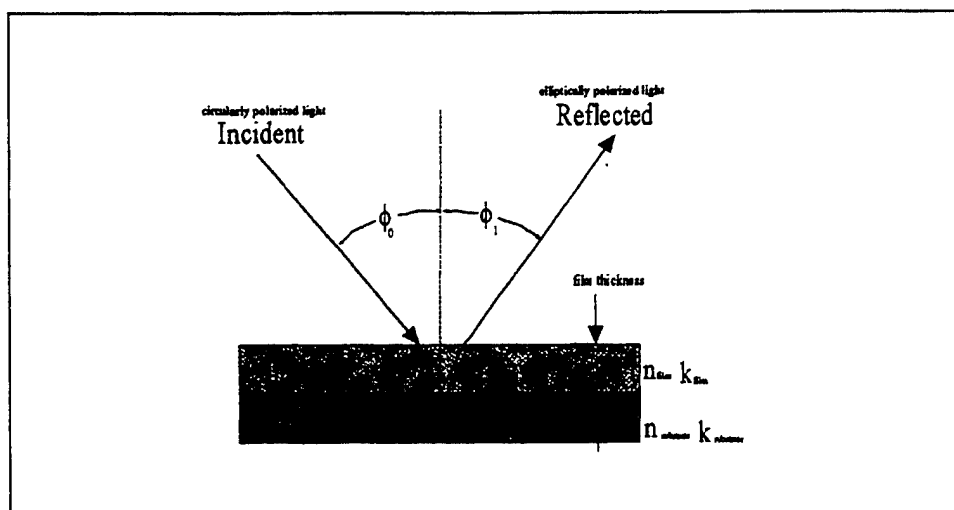


Figure 6 Illustration of Reflection from a Composite Structure.

In the attachments, it is explained why several sets of the ellipsometry parameters are required if the refractive indices and thickness of the film are to be estimated. This requirement has a precise theoretical basis but also helps in coping with noise in the data.

As shown schematically in Figure 7, the task is really quite difficult. Given (say) four sets of values for  $\psi$  and  $\delta$ , a neural net is asked to estimate the asymptotic end point of the  $\psi$ - $\delta$  spiral, which can be readily translated to yield the  $n$  and  $k$  (refractive indices) values of the film.

A simple neural-net architecture for the inversion estimation problem is illustrated in Figure 8.

Extremely accurate inversion can be obtained rapidly with functional-link neural-nets and this approach is feasible especially if the same range of  $(n, k)$  values are encountered from experiment to experiment. Otherwise, the extensive training of the nets can be a burden even with the use of the function-link methodology. This method is not robust

against the occurrence of noise in the ellipsometry readings. More is said about noise in Section 5.

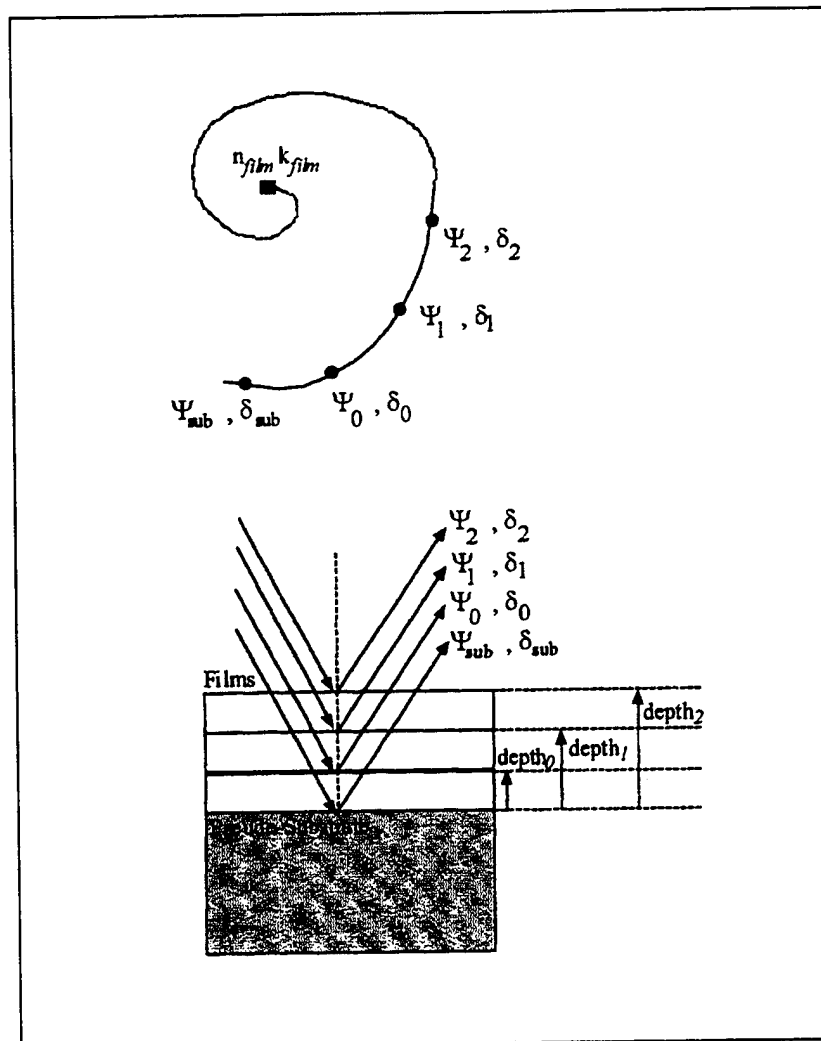


Figure 7 Input and Output Parameters Used in Neural Net.

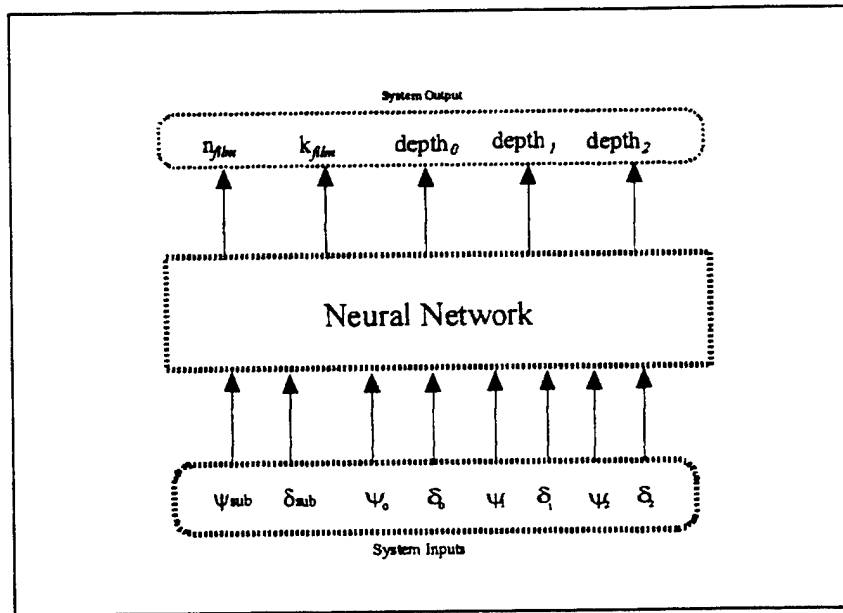


Figure 8 Neural Net for Inverting Fresnel Equations.

### System Identification and Noise Cancellation

Two additional applications are described in Attachment 6. The tasks addressed are those of 'system identification' and 'noise cancellation', both interesting and challenging tasks, made more readily feasible through use of neural net computing, especially with the use of the functional-link net.

In traditional Systems and Control theory, the words 'system identification' usually refer to the task of parameter estimation, estimating the values of the parameters used in the model postulated for the process in question. In contrast to that, in neural-net computing, system identification refers to the task of learning a computational model of the process. In other words, by observing the time dependent input signal and the corresponding time dependent output signal, the neural-net attempts to formulate a computational procedure

which will effect that same transformation, through computation rather than through the 'physical' process. The publication of Attachment 6 describes in detail the considerable success achieved for this type of task for a variety of processes and for a variety of input signal types. This is a capability of use in the control of materials processes.

The work on noise cancellation is very intriguing and deserves further study. The situation is that where a low amplitude signal is drowned in large noise but a measure of that same noise is available through another channel, the second source being free of signal! Although the two noise channels have the same origin, there may be, and there usually are, intervening distorting processes in one or both channels. The task is to use the 'pure' noise channel to predict what the noise should be on the other channel, the one with signal mixed in. One attempts to cancel out the noise, to reduce in fact the output of the difference to zero. One finds that one fails to do so, and the residue, representing the failure, is in fact the signal to be recovered out of the very noisy environment in which it was originally submerged. The results exhibited in Attachment 6 are extremely interesting and are worthy of further examination

## 4. GUIDED EVOLUTIONARY PROGRAMMING

### Overview of Optimization Algorithms

This present program was also able to make a significant useful contribution to the art of optimization. This new method of nonlinear optimization is called guided evolutionary programming with simulated annealing (GESA) and has been of great help in the various applications efforts which this program has interacted with.

Traditional techniques for nonlinear optimization are:

- o Newton-Raphson
- o gradient search
- o conjugate search
- o stochastic search

Common to the first three of these techniques is that they require an analytical description (or an implementation with good numerical accuracy) of the problem. Gradient search and conjugate search have been applied successfully to many, many tasks, even to the task of training neural networks [13]. They are guaranteed to converge to minima, which are likely to be local minima. Conventional stochastic search has the advantage of not getting easily trapped into local minima but the convergence rate is usually very slow because of the lack of guidance in the search.

The GESA algorithm is also a form of stochastic search but benefits from the availability of guidance in its search and is also less prone to being trapped in local minima. In this matter, it borrows and benefits from the practices of *Evolutionary Programming* (EP)

and *Simulated Annealing* (EP). Those two practices were inspired by considerations of processes of nature. The first has its origin in biological evolution and the second is based on considerations of the processes of relieving internal strains induced in solids in cooling or in crystallization.

A brief overview of the paradigms of EP and SA is given in the following to provide a background for description of GESA. Related to all this is the paradigm of *Genetic Algorithms* (GA).

Goldberg [14] has popularized the use of Genetic Algorithms (GA) which are based on the mechanisms of genetics and natural selection. Each solution, described as a parent or child, is coded as a binary vector (a string). Fogel [15] is associated with a paradigm that is called Evolutionary Programming (EP). It is based on Darwin's evolution theory. These two paradigms are basically the same and the basic algorithm is shown in Figure 9

```
generate N number of initial parents
repeat
    generate M children from the parents
    (distributed among parents according to some measure of
    merit of parents)
    evaluate all N+M solutions
    select the N best solutions as parents for next generation
until solution is found
```

Figure 9 A Skeletal Form of The Basic GA/EP Algorithm

The conceptual difference between GA and EP is in the way children are generated. In GA a child is generated by combining two parents (crossover) and then applying a random change (mutation). In EP a child is generated from one parent by a random change. In addition to that, GA has fixed on the idea of representing a solution as a binary



string. This has the disadvantage that the representation is discrete, which means that GA's primary application area is that of combinatorial optimization. Therefore in continuous optimization, a somewhat strained implementation is enforced. It has been argued whether GA or EP is best, in other words if crossover is good or bad. No definitive answer has been given. The question of which one to choose seems to be problem and implementation dependent.

By applying a Monte Carlo simulation procedure [16] to annealing, Kirkpatrick, Gelatt and Vecchi proposed the Simulated Annealing (SA) [17] technique for use in optimization. The algorithm is listed in Figure 10.

```

set initial temperature t
generate randomly a solution
evaluate the solution ->  $y_{best}$ 
repeat
    repeat k(t) times
        generate a new solution from the current best solution
        evaluate the new solution ->  $y_{new}$ 
        accept the new solution as current best solution if
             $\exp[-(y_{new} - y_{best})/t] > p$ 
        decrease t
until solution found

Notations:
t is the temperature
 $y_{new}, y_{best}$  are the objective values of the new and current best solutions
respectively.
p is a random number uniformly distributed between 0 and 1

```

Figure 10 The Basic SA Algorithm

In the form exhibited in that figure, a lower objective value is a better one. The condition for checking if a new solution is going to be accepted as the current best solution is:

$$\text{accept if } \exp[-(y_{new} - y_{best})/t] > p \in [0,1] \quad (4.1)$$

The purpose is to always accept a new solution if it is better than the best current one, and with a probability proportional to how good it is, also accept it even if it is not as good as the current best solution.

The two paradigms, GA/EP and SA are very similar and are the basis of the GESA algorithm.

Before proceeding to describe and explain GESA, a comparison is made of the GA/EP and SA algorithms. A good optimization technique should be guided, it should have the ability to escape from local minima and the ability to converge to a solution with arbitrarily good accuracy. The following comparison is based on these three criteria and the similarities and differences are explained.

- o Regarding ability to escape from local minima.

**GA/EP:** does not have any special mechanism for that but parallelism would decrease the probability of getting trapped in local minima.

**SA:** is not parallel but a trial solution might be accepted even if it is not as good as the current best solution, this is because of the rule given by expression 4.1

- o The issue of how the process is guided so as to generate new trial solutions in the more promising regions.

**GA/EP:** Only the best solutions become parents. To that extent there is some guidance in GA.

**SA:** A better solution has larger probability of becoming the 'parent' for the next generation.

o Regarding the speed and accuracy of convergence.

**GA/EP:** No special mechanism.

**SA:** Convergence to global minimum is assured in principle, but only in an asymptotic sense given infinitely long periods of search.

#### Guided Evolutionary Programming with Simulated Annealing Algorithm

Many algorithms have evolved from the original GA/EP and SA concepts. The GESA algorithm is one of these, perhaps are particularly attractive one. It combines the best characteristics of GA/EP and SA synergistically, by introducing the concept of many different families carrying out searches concurrently in parallel. Comparison of the quality of solutions being obtained by the different families provides guidance to the entire effort of how to allocate search resources to more promising localities. This is not a characteristic which can be duplicated by searching longer. Simulated annealing prevents too early a dismissal of seemingly non-promising localities. The algorithm is well suited for implementation in parallel computers of the SIMD (Single Instruction Multiple Data) type because the total number of children is kept constant.

One form of the GESA algorithm is listed in Figure 11. There can be many slightly different variations of the theme. For example, in the choice of parents, it could be stipulated that the current parent could also compete for the role of being the parent for the next generation. There is also a decision to be made whether families die out or not. The mode of generation and especially the details of how that varies or do not vary with temperature can be specified in slightly different ways.

The performance of GESA has been compared experimentally with those of other algorithms in combinatorial and continuous optimization tasks [18] as well as in resource allocation applications[19]. The conclusion of these benchmark investigations is that it can compete well with GA, EP, SA and Hopfield Net procedures in continuous as well as combinatorial optimization tasks.

Copies of the two cited GESA references are available as Attachments 7 and 8.

#### Applications Including Interpretation of Noisy Ellipsometry Data

Two applications of GESA are mentioned. One being the use of GESA for learning network weights in the training of a multi-layer feedforward neural net. This works perfectly well but might be considered to be less efficient than the functional-link or radial basis function approaches in well structured circumstances. The GESA approach might be the appropriate one for more irregular and more complex net architectures. Some results of studies have been published, one publication being that of Yip, P. P. C. and Y. H. Pao, entitled 'A perfect integration of neural networks and genetic algorithms' published in *Artificial Neural Nets and Genetic Algorithms*, Pearson and Steel (eds.), pp. 88-91, Springer-Verlag, 1995, and also in *Proceedings of the 2nd International*

**main algorithm:**

```

set initial temperatures  $t_1, t_2$  and  $t_3$ 
generate randomly N parents
evaluate these parents
repeat
    for each family do
        generate children* from parent by a random change that is
        proportional in some manner to  $t_3$ 
        evaluate these children
        find the best child
        accept this child as the parent for the next generation if
             $\exp[-(y_{new} - y_{best})/t_1] > p$ 
        find the number of children that will be generated in each family in
        the next generation by calling the subroutine
        decrease the temperature coefficients
until solution found

```

\* the number of children is M the first time

**subroutine:**

```

for each family I do
     $acc_i = 0$ 
    for each child in family  $i$  do
        if  $\exp[-(y_{child} - y_{best})/t_2] > p$ 
            then  $acc_i = acc_i + 1$ 
     $sum\_acc = \sum_i acc_i$ 
    for each family  $i$  do
        the number of children in next generation is
             $M \bullet N \bullet acc_i / sum\_acc$ 

```

**Notation:**

$t_1, t_2$  and  $t_3$  are the temperatures  
 N is the number of children  
 M is the average number of children in each family  
 $y_{new}, y_{best}$  and  $y_{child}$  are objective values (lower is better)  
 $y_{best}$  is the objective value for the globally best solution found so far  
 $acc_i$  is the number of accepted children in family I  
 $sum\_acc$  is the total number of accepted children

Figure 11 The GESA Algorithm

In another interesting and useful application, the GESA algorithm was used to obtain estimates of the optical constants of thin films being grown on a substrate. The experimental conditions are the same as those described in Section 3, and in Attachments 4 and 5, but circumstances are such that there is much noise. Under those conditions, the accuracy of the estimated values for  $n$  and  $k$ , the real and imaginary parts of the refractive index can degrade significantly with increase in noise. It was found that for those circumstances, GESA can be used to great effect.

Overall, the procedure is a sort of non-linear regression, but with a difference. In traditional nonlinear regression, one postulates a model and then tries to find the best set of values for the model parameters so that the process in question is replicated as best as it can be. In the present case, the model is known; the trajectory for the Fresnel parameters is a spiral. The precise form of the spiral depends on the values of the film  $n$  and  $k$ , and of course also on the film thickness. GESA can be used to advantage to determine the optimum values of  $n$  and  $k$ , such that the sum of the squares of the 'errors' in the placement of the data points be at a minimum. In other words, in Figure 12, all the data points would fall a single spiral if there were no noise in the data. Given that there is noise, the task is to determine the 'perpendiculars' from the points to the spiral and find that spiral for which the sum of the squares of the deviations is a minimum. GESA achieves that task well. Some results are exhibited in Table 1 for the noisy data shown in Figure 12.

In Table 1, the values exhibited in each row represent the values of the refractive indices, real and imaginary parts, for the pseudo substrate and for the film material. The values in the column MSE are indicative of the irreducible noise in the measurements as made evident from the distance of the data points from the best spiral. All values were obtained

with use of GESA. Seven consecutive data points were used for each estimation operation.

This method is the only successful single wavelength technique capable of yielding accurate estimates of optical constants for noisy data.

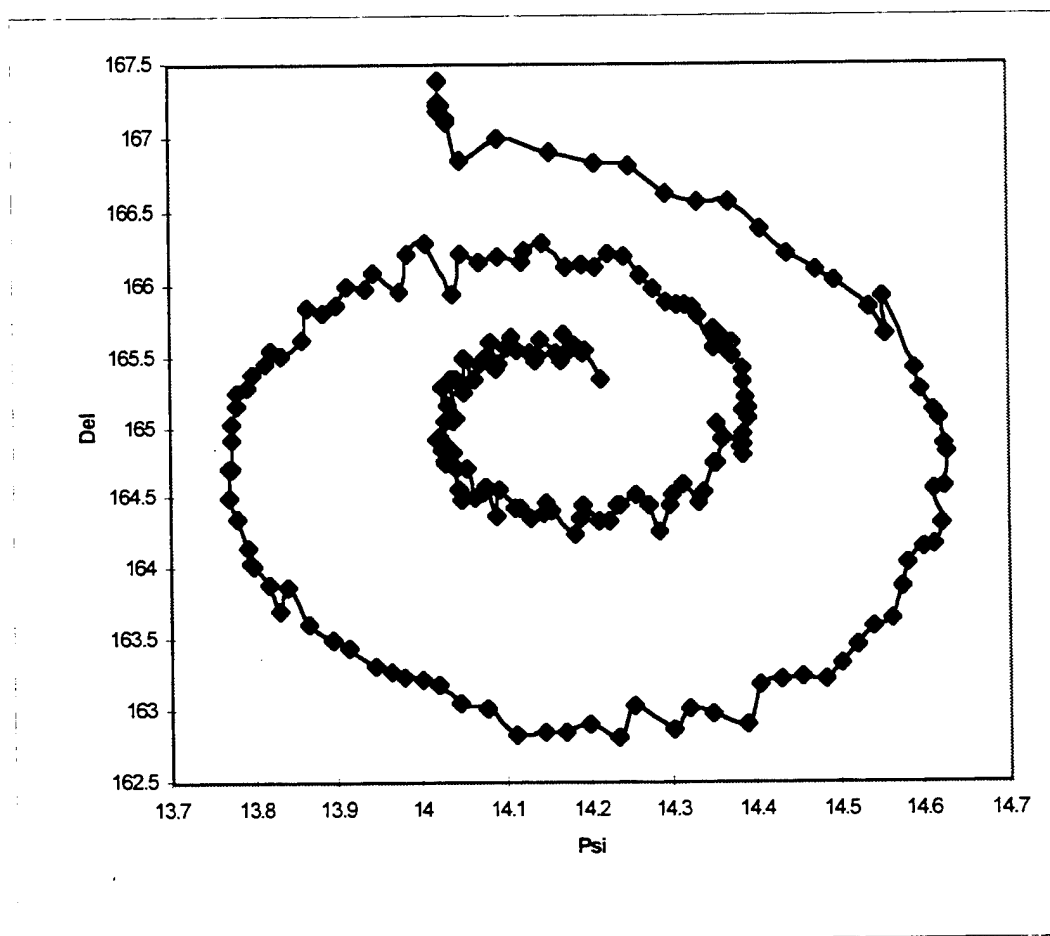


Figure 12 Noisy Ellipsometry Data.

Table 1 Estimates of Pseudosubstrate and Film Optical Constants

GESA(50 iterations, 10 families, 10 children)				
ns	ks	nf	kf	MSE
4.308479	0.479483	4.289525	0.501549	0.004215
4.317444	0.500452	4.281033	0.606694	0.006013
4.351112	0.531003	4.254094	0.6229	0.002024
4.363045	0.571209	4.240501	0.612749	0.014641
4.358186	0.617562	4.284027	0.602418	0.002722
4.330631	0.65044	4.294611	0.579446	0.003056
4.293939	0.653908	4.281381	0.602237	0.008632
4.266227	0.648776	4.305117	0.564677	0.004647
4.180917	0.582457	4.162605	0.827471	0.085035
4.219191	0.582555	4.26225	0.611928	0.011766
4.218979	0.551111	4.275157	0.639098	0.007338
4.22225	0.519464	4.235114	0.61458	0.028134
4.260075	0.519567	4.281833	0.592483	0.006377
4.29939	0.523912	4.306261	0.541538	0.012133
4.31369	0.527614	4.299119	0.57799	0.003927
4.322211	0.542275	4.283049	0.592842	0.003949
4.333405	0.557471	4.2865	0.594473	0.003409
4.343277	0.579774	4.29328	0.526521	0.003224
4.327979	0.583742	4.264687	0.564518	0.007897
4.315785	0.602803	4.32217	0.534342	0.008935
mean		4.2751157	0.5955227	0.0114037
std dev		0.034137894	0.065026582	0.018327729



## **5. APPLICATION OF FUNCTIONAL-LINK AND GUIDED EVOLUTIONARY PROGRAMMING TO MATERIALS FORMULATION**

The activities of this research program were carried out in close co-ordination with those of the Materials Process and Design program of Wright Laboratory, and as a result there were many opportunities for testing its methodologies on real tasks, to assess the validity of its results and the efficiency of its procedures.

As a result of interactions with the Materials Directorate and with Case Western Reserve University, a neural-net computing company has developed a powerful and easy-to-use computer software package for optimal formulation of material compositions. The CAD/Chem [trademark of AI Ware, Inc.] system utilizes the Functional-link and GESA paradigms to help materials researchers design new material compositions in optimal manner.

The software system, CAD/Chem, is an adaptive intelligent system which acquires instances of material composition and corresponding property values, and synthesizes a computational model of the material. Subsequently, if a specific set of property values is desired, CAD/Chem is able to suggest formulations which would meet the desired goal.

In most cases, the desired property values are not precise or 'crisp' values but can vary over a range of values to different degrees of acceptability. That aspect is accommodated though the use of desirability functions, in a manner similar to the use of membership functions in Fuzzy Sets. In addition not all properties are of equal importance, and so the goal values can be weighted.

Finally when appropriate the new material can be specified in an optimal manner, in the sense that, if necessary, a least cost formulation could be specified. Other equality and inequality constraints can also be met in the search for the optimal composition.

The system architecture of CAD/Chem is modular. The CAD/Chem system is truly a versatile modeling and optimization system suitable for use in support of a variety of tasks in materials research including design of experiments, parameter interpretation, sensor validation, materials formulation and so on. It is supported by good graphics and input/output capabilities, but it owes its power principally to the efficiency of the Functional-link and GESA paradigms.

Through the co-operation of AI Ware Inc., a demonstration version of CAD/Chem 4.5 is made available in diskette form, for illustration purposes, together with an accompanying manual as Attachment 8.

## **6. APPLICATION OF FUNCTIONAL-LINK AND GUIDED EVOLUTIONARY PROGRAMMING TO OPTIMAL CONTROL OF NONLINEAR SYSTEMS**

Neural networks are ideally suited to the tasks of monitoring and control of nonlinear dynamic systems especially systems which are 'opaque' in the sense that the behavior cannot be described analytically nor modeled simply in some linear manner.

There are three distinctly different approaches to the task of control. These might be called:

- o the 'Backpropagation in time' method,
- o the Inverse Net method, and
- o the Optimal Control method.

The Backpropagation-in-time method is illustrated schematically in Figure 13. In this approach a neural-net model is learned of the system. By this statement, it is meant that given the state of the system and the current control action, the neural-net will compute the value of the next state. The word 'state' is used in the conventional sense and may entail a number of time-lagged values of some vector quantity described the current descriptors of the system.

In control mode, the desired value of the next state is specified and the value of the requisite control equation is computed quite simply in a Backpropagation of error manner. In control mode all network parameter values are already known having been fixed during the training stage.

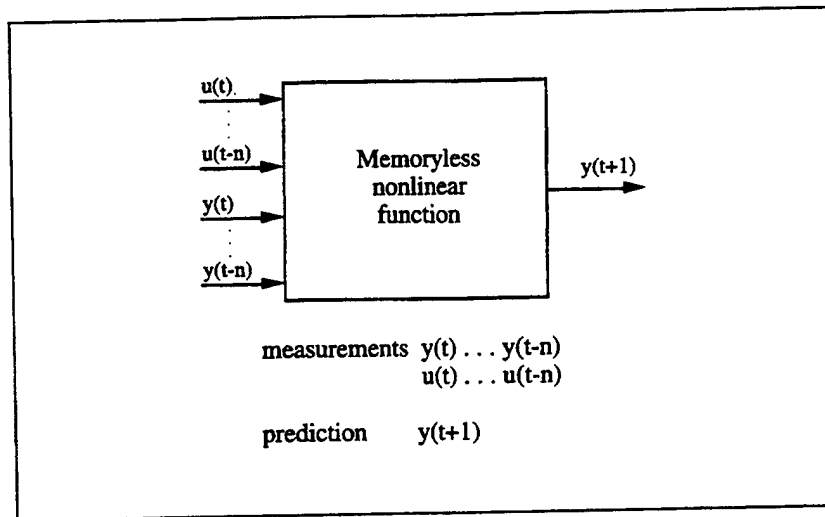


Figure 13 The Modeller/Predictor Neural Net.

The inverse net approach is illustrated schematically in Figure 14. The same data used in the learning of a model of the system could also be used to learn an 'inverse' net shown in Figure 14. In this case the inputs are the current system state and the desired value of the next state, and the net computes the requisite value of the control action. This is clearly feasible but there are cases where the inversion is not unique. If many control actions can result in nearly the same value of the next state, then specifying the desired value of the next state might lead to an average of the several possible control actions, with the average being an incorrect solution.

The third approach is one that is championed by this research program and is made feasible primarily because of the availability of the Functional-link and GESA. It is different from the previous two methods in that a trajectory with some particular attributes can be specified, and a sequence of control actions which would produce such a trajectory in an optimal manner is given as the solution.

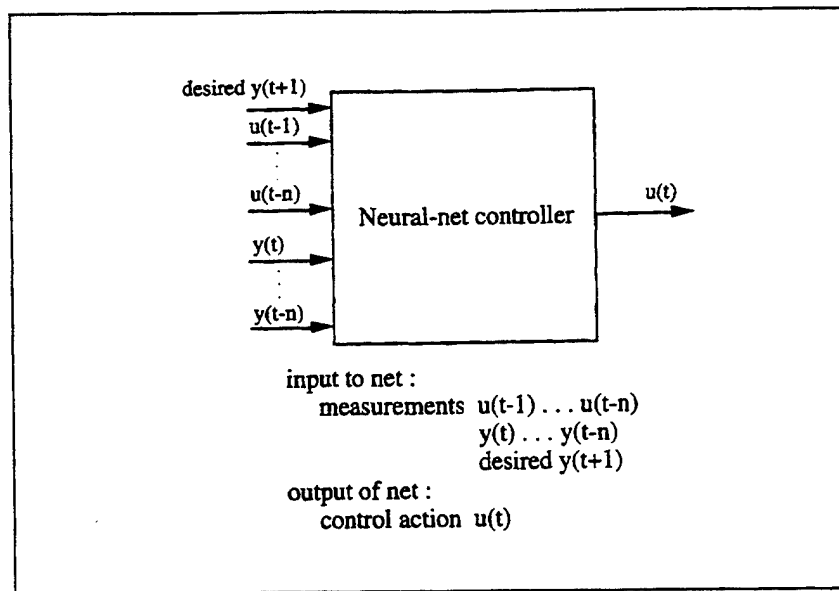


Figure 14 The Neural Control-Action Generator.

A detailed discussion of this topic is given in Attachment 10, a publication by members of this research team [20].

GESA may also be used for automatic formulation of optimal fuzzy control. This is a subject matter that is not directly related to that of this section but is related nevertheless to the task of intelligent controls. That topic is described in Attachment 11, also a publication which originated from work carried out in this program [21]

There also exists a commercially available software product called the Process Advisor [trademark of AI Ware, Inc.] which models dynamic systems, and a product called Neusight [trademark of Pegasus Inc.] which applies such technology to the optimal control of systems. Both of these products are based on the functional link and GESA.

## 7. THE ROLE OF SELF-ORGANIZATION IN DATA ANALYSIS

In the analysis and management of multidimensional data, ultimately the most basic task is that of self-organization. In self-organization, the data points need to evolve a way of describing the manner in which they are distributed in the multidimensional space of the data points. This is not an easy task because the density distribution may be very complex, topologically speaking.

A conventional way of proceeding is to attempt to form clusters of data points, and hierarchies of clusters. Often such clusters might also be categories, in the sense that all data points in any one cluster also have certain characteristics in common. In any event, with use of clusters and hierarchies of clusters, a very large number of data points (or patterns) can be stored and retrieved efficiently. This can be the basis for use of such data structures as associative memories. After self-organization, all the data points, or objects, or equivalently, patterns, are in one cluster or other. One of the attributes of a cluster is therefore the list of the identifiers of all its member patterns. Other attributes are the characteristics common to all members of the cluster, perhaps to varying degrees. Any new pattern is readily determined to be closer to one of the existing clusters, more so and to any of the other clusters. If it is 'within' the bounds of that cluster then some prediction can be made regarding its characteristics, if not, then it might be the basis for a new cluster, with new characteristics.

In one of the efforts of this program, in the development of the Rapid Foundry Tooling System, there was opportunity to assess the feasibility of using self-organization to evolve a system architecture which could be used for trouble shooting in the practice of metal casting.

Some troubleshooting capability can be demonstrated using the diskette made available as Attachment 12. Using that diskette it can be demonstrated that any new casting with defects can be recognized to be similar to one or more cases experienced previously. In other words, a faulty casting can be 'recognized' to be so in a rather broad manner, the probable causes are identified and so are some of the likely 'cures'. Such matters can also be implemented with static links in some hypertext manner, but that approach would not be able to support the same type of robust and efficient search and retrieval capability provided by the associative memory.

## **8. REDUCED DIMENSION REPRESENTATION OF DATA: THE CONSTRAINED VARIANCE APPROACH.**

This section continues the line of thought of the previous section but in a different vein.

It is difficult to make sense out of a large body of multi-featured pattern data. Actually the body of data need not be large; even a set of 400 patterns each of six features would be quite difficult to 'understand'. The idea of self-organization has to do with that type of situation and can be understood in terms of two main approaches to that task. In the one case, endeavor is directed to discovering how the data are distributed in pattern space, with the intent of describing large bodies of patterns more simply in terms of multi-dimensional clusters or in terms of some other distributions, as appropriate. This is the dominant concern underlying the ART[22], ISODATA [23] and feature map [24] approaches.

In the other case, effort is devoted to dimension reduction. The idea is that perhaps the original representation with a large number of features is redundant in its representation, with several features being near repetitions of each other; in which case principal feature extraction accompanied by dimension reduction would simplify the description of each and all the patterns. Clustering could be subsequently achieved in the reduced dimension space. The Karhunen-Loeve (K-L) transform[25], neural-net implementations of the K-L transform[26], and the auto-associative memory [27] are all directed to principal component analysis (PCA), feature extraction and dimension reduction.

Actually the two streams of activity are not entirely independent. For example, the ART approach has a strong winner-take-all mechanism in forming the clusters. It can be viewed as 'extracting' the principal prototypes, and forming a reduced dimension



description in terms of a few category prototypes. Similarly, the feature map approach aims at collecting similar patterns together through lateral excitation-inhibition so that patterns with similar features are mapped into contiguous regions in a reduced dimension feature map. That method clusters and reduces dimensions also.

The work of this program gave rise to efforts aimed at principal component analysis, in nonlinear manners, so as to reduce dimensions and perhaps in that way reveal what was important in any description of material composition or material process. Those efforts induced the development of a new approach to the task of self-organization. The idea is that data be subjected to a nonlinear mapping from the original representation to one of reduced dimensions. The mapping is implemented with a multilayer feedforward neural net. The parameters of the net are learned in an unsupervised manner based on the principle of conservation of the total variance in the description of the patterns.

The concept of dimension reduction is strange in itself. In what way can a reduced-dimension description of a body of pattern data be representative of the original body of data? The answer is known for the linear case but is more difficult to detail in the general nonlinear case. Instead, in the present discussion, the approach is simply described in terms of conservation of variance in connection with a nonlinear transformation, and the consequences of such mapping are examined for some bodies of data in Attachment 12 [28].

First of all, this method was applied to a body of data of bench-mark standing, regarding the quality of various gasoline blends. In the reduced dimension representation, the result of this new type of mapping yielded a 2D display similar to what might be expected from a feature map mapping, in an interesting manner. Patterns which have similar research octane ratings are mapped automatically into contiguous regions in the 2D reduced

dimension mapping. There is no formation of clusters. Instead a rather general spread out measure of similarity and associated correspondence in octane rating can be visualized. It becomes clear that high octane rating can be realized in manner ways and there is guidance towards the formulation of improved blends.

Application of the method to complex sensor data indicated, once again, that patterns representing 'fault' conditions became self-organized into contiguous regions, albeit of rather free form, in 2D, distinct from the patterns representing 'no-fault'.

In both of those two mappings, the category or property value must have been associated strongly with the pattern descriptions. The reduced-dimension mapping merely made that circumstance more obvious and more easily visualized. In yet another case the same approach was applied to a sparse body of data, sparse in the sense of not having many exemplars but also sparse in the sense that many feature values were missing so that in fact only a small subset of features were available for this exercise. The data were for a body of crystal structure parameters for semiconductors and there was interest in seeing whether certain regions of crystal structure 'space' was associated with low bandgaps. The reduced 2D map did give hints as to what regions might be fruitful for further exploration.

All these matters are described in some detail in Attachment 12.

## 9. SUMMARIZING REMARKS

A very large amount of work was carried out under the sponsorship of Air Force Contract No. F33615-87-C-5250. That work was carried out at Case Western Reserve University, at Wright Laboratory and also at various facilities and companies which were collaborating in research and development with Wright Laboratory; these latter sites included Kelley Air Force Base in San Antonio, TX, and AI Ware Inc. in Cleveland.

The contract was from February, 1988 to December 1995, but this report covers only the period from October 1992 to December 1995. Previous work had been reported in the Interim Report WL-TR-934021.

Even so only a portion of the work and results are addressed by this report, and then mostly through the device of referring the reader to reprints or preprints of papers authored by researchers of this program, publications in established technical journals of archival quality. The report is divided into two volumes, for the convenience of the reader, the second volume being the collection of the attached publications and two sets of diskettes. It has not been possible to include all the technical journal papers published. It has not been possible to address all the work done either. Otherwise the process of report preparation would have been stupefying and a multi-volume report would have been mind-numbing.

As it is, this report is happy to conclude with the thought that this effort helped to launch four vigorous original streams of innovation into the practice of adaptive, parallel, distributed computing, these four being the functional-link net, guided evolutionary programming with simulated annealing (GESA), the self-organizing associative memory

and the constant variance mapping for 2D displays. These methods have found applications in materials research and have given rise to significant commercial products.

## 10. REFERENCES

- [1] Rumelhart, D. E., J. L. McClelland and the PDP Research Group (Eds.), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. Vol. 1: *Foundations*, MIT Press, Cambridge, MA, 1986.
- [2] Pao, Y. H., *Adaptive Pattern Recognition and Neural Networks*, Addison-Wesley, Reading, MA, 1989.
- [3] Pao, Y. H., G. H. Park, and D. J. Sobajic, Learning and generalization characteristics of the random vector Functional-Link net, *Neurocomputing*, vol. 6, no. 2, Elsevier Publishing, Amsterdam, 1994, pp. 163-180.
- [4] Igel'nik, B. and Y. H. Pao, Stochastic choice of basis functions in adaptive function approximation and the Functional-Link net, *IEEE Transactions on Neural Networks*, vol. 6, 1995, pp. 1320-1329.
- [5] Marks, R. J. II, *Introduction to Shannon Sampling and Interpolation Theory*, Springer-Verlag, New York, 1991.
- [6] Traub, J. F. and H. Wozniakowsky, Breaking intractability, *Scientific American*, January 1994, pp. 102-107.
- [7] Kolmogorov, A. N., On the representation of continuous functions of several variables by superposition of continuous functions of one variable and addition, *Doklady Akademii. Nauk SSSR*, vol. 114, 1957, pp. 679-681.
- [8] Sprecher, D. A., On the structure of continuous functions of several variables, *Transactions of the American Mathematical Society*, vol. 115, 1964, pp. 340-355.
- [9] Hecht-Nielsen, R., Kolmogorov's mapping neural network existence theorem, *Proc. IEEE First International Conference on Neural Networks*, SOS Printing, San Diego, CA. 1987, vol. II, pp. 11-14.
- [10] Sprecher, D. A., A universal mapping for Kolmogorov's superposition theorem, *Neural Networks*, vol. 6, 1994, pp. 1089-1094.
- [11] Yao, T., *The Technology and Physics of Molecular Beam Epitaxy*, Plenum, New York, 1985.
- [12] Azzam, R. M. A., and N. M. Bashara, *Ellipsometry and Polarized Light*, North-Holland, New York, 1977.
- [13] McCrackin, F. L., E. Panaglia, R. Stromberg, and H. S. Steinberg, Measurement of the thickness and refractive index of very thin films and the optical properties of surfaces by ellipsometry, *Journal of Research of NBS*, Vol. 67A, 1963, pp. 363-377.
- [14] Goldberg, D., *Genetic algorithms in search, optimization, and machine learning*, Addison-Wesley, Reading, MA, 1989.
- [15] Fogel, L. J., A. J. Owens, and M. J. Walsh, *Artificial intelligences through simulated evolution*, Wiley, New York, 1966.

- [16] Metropolis, N. et al, Equation of state calculations by fast computing machines, *Journal of Chemical Physics*, Vol. 21, 1953, pp. 1087-1092.
- [17] Kirkpatrick, C. D. Gelatt Jr. and M. P. Vecchi, Optimization by simulated annealing, *Science*, Vol. 220, 1983, pp. 671-680.
- [18] Yip, P. P. C. and Y. H. Pao, Combinatorial optimization with use of guided evolutionary simulated annealing, *IEEE Transactions on Neural Networks*, Vol. 6, 1995, pp. 290-295.
- [19] Yip, P. P. C. and Y. H. Pao, A guided evolutionary computation technique as function optimizer, *Proceedings of the First IEEE Conference on Evolutionary Computation*, Orlando, FL, 1994, pp. 628-633.
- [20] Nyberg, M. and Y. H. Pao, Automatic optimal design of fuzzy systems based on universal approximation and evolutionary programming, *Fuzzy Logic and Intelligent Systems*, H. L. Hua and M. Gupta (Eds.), Chapter 12, Kluwer Academic Publishers, Norwell, MA, 1995.
- [21] Carpenter, G. A. and S. Grossberg, "ART2: Self-Organization of Stable Category Recognition Codes for Analog Input Patterns," *Applied Optics*, Vol. 26, 1987, pp. 4919-4930.
- [22] Ball, G. H. and D. J. Hall, ISODATA: an iterative method of multivariate data analysis and pattern classification, *Proceedings of the IEEE International Communications Conference*, Philadelphia, PA, June, 1966.
- [23] Kohonen, T., *Self-Organizing Maps*, Springer-Verlag, New York, 1995.
- [24] Fukunaga, K. and W. L. G. Koontz, Application of the Karhunen-Loeve expansion to feature selection and ordering. *IEEE Transactions on Computers*, Vol. 19, 1970, pp. 311-318.
- [25] Oja, E., A simplified neuron model as a principal component analyzer. *Journal of Mathematics and Biology*, Vol. 15, 1982, pp. 267-273.
- [26] Kramer, M., Nonlinear principal component analysis using autoassociative feedforward neural networks. *AIChE*, Vol. 37, 1991, pp. 233-243.
- [27] Pao, Y. H. and C. Y. Shen, Self-organization of pattern data with dimension reduction through learning of non-linear variance-constrained mapping, submitted for publication in *Pattern Recognition*, 1996.

## 11. ATTACHMENTS: REPRINTS, PREPRINTS AND DISKETTES.

Attachment 1. A Historical Perspective on Function Approximation with Neural-Net Computing. [preprint of 1996 paper by Yoh-Han Pao, used to-date as handout in support of tutorials and conference plenary talks].

Attachment 2. Learning and Generalization Characteristics of the Random Vector Functional-Link Net. [reprint of 1994 paper by Yoh-Han Pao, G. H. Park and D. J. Sobajic, published in *Neurocomputing*, Vol. 6, pp. 163-180].

Attachment 3. Stochastic choice of basis functions in adaptive function approximation and the Functional-Link net [reprint of 1995 paper by B. Igel'nik and Y. H. Pao, published in *IEEE Transactions on Neural Networks*, Vol. 6, pp. 1320-1329].

Attachment 4. Neural-net based optical ellipsometry for monitoring growth of semiconductor films [reprint of paper by G. H. Park, Y. H. Pao, K. G. Eyink, S. R. LeClair and M. S. Soclof, presented at 1994 IFAC *International Symposium on Artificial Intelligence in Real Time Control*, October 3-5, Valencia, Spain, and published in the Proceedings of the Symposium].

Attachment 5. Neural-net computing for interpretation of semiconductor film optical ellipsometry parameters [preprint of a 1996 paper by G. H. Park, Y. H. Pao, B. Igel'nik, K. G. Eyink and S. R. LeClair, *IEEE Transactions on Neural Networks* (accepted for publication, in press)].

Attachment 6. System identification and Noise Cancellation via Neural-Net Computing [reprint of 1994 invited paper by G-H. Park and Y. H. Pao presented at the IEEE World Congress on Computational Intelligence, Orlando, Florida, June, 1994, and published in the Proceedings of that Congress].

Attachment 7. Combinatorial optimization with use of guided evolutionary simulated annealing [reprint of 1994 paper by P. P. C. Yip and Y. H. Pao published in *IEEE Transactions on Neural Networks*, vol. 6, pp. 290-295].

Attachment 8. A guided evolutionary simulated annealing approach to the quadratic assignment problem [reprint of 1994 paper by P. P. C. Yip and Y. H. Pao published in *IEEE Transactions on Systems, Man and Cybernetics*, vol. 24, pp. 1383-1387].

Attachment 9. Demonstration version of the CAD/Chem software system, commercial software product of AI Ware Inc., for optimal product formulation and design of experiments [2 Diskettes].

Attachment 10. The functional link net and learning optimal control [reprint of a 1995 paper by Y. H. Pao and S. M. Phillips, published in *Neurocomputing*, vol. 9, pp. 149-164].

Attachment 11. Automatic optimal design of fuzzy systems based on universal approximation and evolutionary programming [reprint of a 1995 paper by M. Nyberg and Y. H. Pao, published as Chapter 12 in *Fuzzy Logic and Intelligent Systems*, edited by H. L. Hua and M. Gupta, Kluwer Academic Publishers, Norwell, MA].

Attachment 12. Diskette of code demonstrating use of self-organized associative memory for trouble shooting of the design of metal cast parts [work of M. Soclof and David Yan].

Attachment 13. Self-organization of Pattern Data with Dimension Reduction Through Learning of Non-Linear Variance-Constrained Mapping [a 1996 manuscript by Y. H. Pao and C. Y. Shen, submitted for publication in *Pattern Recognition* ].

Attachment 14. Title pages of U.S. Patent and U.S. Patent application inspired by results of work of this research effort.